1-1-1999

A New Technique for Estimating Confidence Intervals on DEA Efficiency Estimates

Janice Forrester
Cytera Systems, Inc

Timothy R. Anderson
Portland State University, tim.anderson@pdx.edu

Citation Details

Forrester, J.; Anderson, T.R., A New Technique for Estimating Confidence Intervals on DEA Efficiency Estimates. Management of Engineering and Technology, 1999. Technology and Innovation Management. PICMET '99. Portland International Conference on, 1999.

A New Technique for Estimating Confidence Intervals on DEA Efficiency Estimates

Janice Forrester[1], Timothy R. Anderson[2], Ph.D.
[1]Director of Strategic and Analytical Services, Cytera Systems, Inc.
9220 SW Barbur Blvd., Suite 119-148, Portland, OR 97213
[2]Assistant Professor, Engineering Management Program, Portland State University
PO Box 751, Portland, OR 97207 - 0751

Since the initial development by Charnes, Cooper, Rhodes in 1978, Data Envelopment Analysis (DEA) has been widely used as a tool for examining the efficiency of sets of decision making units, DMUs in a wide variety of application areas, such as agriculture, manufacturing, education, and health care. By its very nature, DEA is a deterministic tool that allows direct comparison of each DMUs with only a small set of additional assumptions. The formulation of the traditional input oriented DEA model is given by:

$$\begin{aligned} &\min_{\theta,\,\lambda} \theta, \\ &s.t.\ \ Y\lambda \geq y_0, \\ &\qquad X\lambda - \theta x_0 \leq 0, \\ &\qquad \lambda \geq 0, \theta \text{ free.} \end{aligned} \tag{1}$$

However, a number of examples can be easily found to show that the effort (input) indices and the performance (output) indices are commonly of a stochastic nature. That is, real world production relationships are often subject to stochastic effects, herein lies a major criticism of a deterministic DEA study. To respond to this problem, chance constrained formulations of the DEA problem have been recently developed by various individuals, which attempt to consider stochastic variations in data.

Chance constraints are designed to consider the fact that the data is only a sample of the total population. Therefore the production frontier that is constructed from the sample is not necessarily the true average production function but rather a best guess. It is the purpose of this project to develop and demonstrate a methodology for calculating a confidence band for the estimated production function so that we can specify, with a predetermined level of confidence, an interval containing the most likely efficiency score for each DMU considered.

To date, the primary school of thought on how to incorporate stochastic variation into DEA models involves nonlinear optimization proposed by Land, Lovell and Thore (1993) as well as Olesen and Petersen (1995). The input oriented DEA chance constraint model, assuming only the outputs are stochastic, is of the following form:

$$\begin{aligned} &\min_{\theta,\,\lambda} \theta, \\ &s.t.\ \ \mathrm{E}(Y\lambda - y_0) \pm c_{ritical}\sqrt{\lambda^2 Var(Y)} \geq 0, \\ &\qquad X\lambda - \theta x_0 \leq 0, \\ &\qquad \lambda \geq 0, \quad \theta \text{ free.} \end{aligned} \tag{2}$$

The chance constraints presented in formulation (2) attempt to integrate a stochastic methodology with that of linear optimization. The nonlinear system of equations fails to consider that the probability is constrained by the system and that the constraints on the probability impact the system in which it is embedded. Thus, it leads to a violation of the convexity assumption and biasing of the lambda values.

It is our intention to determine, on average, what the production function would be, and given that we do not know the true mean value, we will need to estimate a confidence band that traps the true mean production function. We are interested in the average production since we assume that the individual sample points will vary due to noise, in which case the average is an appropriate measure of central tendency. We have developed a two-phase approach that separates the deterministic estimation of the production function from the stochastic variation in the data. Phase I uses the extreme point method of DEA to determine the expected production function by solving for the mean efficiency score, $\bar{\theta}$. Phase II for DMU_0 compares DMU_0 to all other DMUs adjusted for variability due to noise. Based on (2), we see that the variability due to noise is symmetrical about the mean given by plus/minus some standard deviation. The standard deviation can be estimated in the traditional fashion directly from the data. As a result we are able to determine an upper and lower confidence band on the production function, which in turn can be used to place bounds on the efficiency score for each DMU. In addition, our chance constrained model is applicable to any defined parametric distribution, easily adapted to consider covariance, can be solved using linear programming rather than nonlinear programming, and does not violate the DEA convexity assumption.